Timothy Filippone
CS432
Assignment 3

Problem 1

To retrieve html files this command was used:

```
wget -i urlist.txt
```

To extract text from a given URI, getText() from A3.py was used. It calls a request and uses JusText to remove boilerplates.

Problem 2

To check matches this command was used:

```
grep -rnw /path/to/text/ -e 'query_term' | wc -w
```

TFIDF	TF	IDF	URI
0.408182	0.02	20.4091	http://bit.ly/2ENbqDM
0.92861405	0.0455	20.4091	http://dailycaller.com/2018/02/18/trump-oprah-winfrey-60-minutes/?utm_source=site-share
0.43471383	0.0213	20.4091	http://read.bi/2jWZ4xP
0.463898843	0.02273	20.4091	http://politi.co/2ECvmtV
0.31430014	0.0154	20.4091	http://politi.co/2FcZxoj
0.40001836	0.0196	20.4091	https://fb.me/Ny5nBy4z
0.37960926	0.0186	20.4091	https://en.pakistantv.tv/2018/02/19/trump-faces-calls-to-act-against-russia-after-muellers-indictments/
0.32858651	0.0161	20.4091	http://bit.ly/2ofRTI5
0.47553203	0.0233	20.4091	http://cnn.it/2jAtPcK
0.30001377	0.0147	20.4091	https://fb.me/7vZpAwBH2

Timothy Filippone
CS432
Assignment 3

Problem 3 results:

Using pr.domaineye	
PageRank	URI
8	http://bit.ly/2ENbqDM
6	http://dailycaller.com/2018/02/18/trump-oprah-winfrey-60-minutes/?utm_source=site-share
7	http://read.bi/2jWZ4xP
7	http://politi.co/2ECvmtV
7	http://politi.co/2FcZxoj
8	https://fb.me/Ny5nBy4z
2	https://en.pakistantv.tv/2018/02/19/trump-faces-calls-to-act-against-russia-after-muellers-indictments/
N/A	http://bit.ly/2ofRTl5
9	http://cnn.it/2jAtPcK
N/A	https://fb.me/7vZpAwBH2